Industrial and Manufacturing Systems Engineering Conference Proceedings and Posters

Industrial and Manufacturing Systems Engineering

9-28-2017

Operationalizing the C's of Teamwork in an Intelligent Tutoring System

Desmond Bonner
Iowa State University, dbonner@iastate.edu

Kaitlyn Ouverson
Iowa State University

Stephen B. Gilbert
Iowa State University, gilbert@iastate.edu

Michael C. Dorneich
Iowa State University, dorneich@iastate.edu

Eliot H. Winer
Iowa State University, ewiner@iastate.edu

See next page for additional authors

Operationalizing the C’s of Teamwork in an Intelligent Tutoring System

Abstract

One of the difficulties in creating a team-focused intelligent tutoring system (ITS) is defining the measures used to assess the team’s performance. While the team research literature offers nine C’s of teamwork to consider, e.g., cooperation, communication, etc., it can also be difficult to implement these in real-world practice. This paper reviews the approach used in three team ITSs in which the C’s were used, offering guidance for future implementation of team tutors.

Disciplines

Ergonomics | Human Factors Psychology

Comments

This proceeding is published as Bonner, Desmond, Kaitlyn Ouverson, Stephen Gilbert, Michael Dorneich, Eliot Winer, Anne Sinatra, Anastacia Macallister, and Adam Kohl. "Operationalizing the C’s of Teamwork in an Intelligent Tutoring System." In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61, no. 1 (2017): 745-749. DOI: 10.1177%2F1541931213601671.

Authors

Desmond Bonner, Kaitlyn Ouverson, Stephen B. Gilbert, Michael C. Dorneich, Eliot H. Winer, Anne Sinatra, Anastacia Macallister, and Adam Kohl

Operationalizing the C's of Teamwork in an Intelligent Tutoring System

Desmond Bonner[1], Kaitlyn Ouverson[1], Stephen Gilbert[1], Michael Dorneich[1],
Eliot Winer[1], Anne Sinatra[2], Anastacia Macallister[1], Adam Kohl[1]
[1]Iowa State University
[2]U.S. Army Research Laboratory

One of the difficulties in creating a team-focused intelligent tutoring system (ITS) is defining the measures used to assess the team's performance. While the team research literature offers nine C's of teamwork to consider, e.g., cooperation, communication, etc., it can also be difficult to implement these in real-world practice. This paper reviews the approach used in three team ITSs in which the C's were used, offering guidance for future implementation of team tutors.

INTRODUCTION

When conducting a traditional in-person team experiment, there are many challenges and decisions that need to be made by the researchers. In the case of creating computer-based and intelligent tutoring system (ITS)-based team research studies, the challenges increase even more.

The development of an ITS is a detailed and time-intensive process, requiring often 50-100 hours of development time per hour of training experience (Aleven, McLaren, Sewall, & Koedinger, 2009). Using a domain-independent ITS framework, such as the Generalized Intelligent Framework for Tutoring (GIFT), can help to improve consistency and promote reusability of materials (Sottilare, Brawner, Goldberg, & Holden, 2012). However, the method of creating a self-contained adaptive training experience that adjusts to the individual's characteristics and performance becomes exponentially more complex with a team. The tutor must track what each team member is doing, as well as the overall team performance. In addition, it is important to track not only the performance on the task, but also the team skills – How well does this team work together? To carry out these assessments and evaluations at these multiple levels, specific, research-based measures must be chosen. However, available research on team-based evaluation measures is often quite generalized, in an effort to be broadly applicable, which leaves a chasm of implementation to cross when building an actual tutor. This paper offers guidance to bridging that chasm.

Challenges in both technological implementation and experimental design arise when doing team ITS research. Technologically, an ITS needs to be able to observe its learner, just as a human tutor observes the struggles of a learner, and must adapt feedback accordingly. After developing a learning experience in a particular software platform, the first step of Team ITS research is synchronizing the ITS and the external software environment such that ITS receives the necessary messages from the learning environment and acts on them in real time. Next is implementing decision rules to choose when to give feedback, how much to give, and whether to give it to individuals or the whole team. These rules must combine real-time information about each individual learner with that of the team, ideally using research-based evidence from the learning sciences.

Once the team tutor is constructed, designing an experiment to evaluate the effectiveness of the tutor can also be challenging. Because many variables affect a team's performance, it is difficult to assess after a team learning experience whether the largest factor affecting performance was, for example, the team members' familiarity with each other, the fact that they had performed this task now give times, or the individual team members' agility with the tools of the task.

The C's of teamwork (Salas, Suffler, Thayer, Bedwell, & Lazzara, 2015) provide nine critical considerations of team behavior and teamwork which include Cognition, Cooperation, Conflict, and Communication. These C's can assist in identifying areas in which measures of teamwork can be developed for the conduction of team studies in ITS contexts. Additionally, an important step in developing measures is finding ways to quantify them so that a computer can evaluate them. Sottilare and colleagues (Sottilare et al., in press) developed a series of behavioral markers that can be used to assess team behaviors. In this paper, we describe three case studies of ITS development that have utilized the C's of teamwork (Salas, et al., 2015), and behavioral markers (Sottilare, et al., in press) to demonstrate the approaches and challenges to conducting and developing team ITS research.

Note that the next stage in developing an ITS for teams is creating feedback for the team members, and creating rules for deciding when feedback goes to individuals vs. the entire team. While the authors have explored this issue in Walton et al. (2014) and Bonner et al. (2016), it is beyond the scope of the current paper.

BACKGROUND

The Nine C's

The nine C's fall into two categories: influencing conditions and core processes/emergent states. Salas et al. (2015) stress that there is no order to these considerations and that all may serve as the most important critical consideration, depending on the context. The influencing conditions include the concepts of Context, Composition and Culture, which affect the team's performance somewhat separately from the team's actual task. Team Context is the environment in which the team exists, while Composition is defined by the characteristics of the team members, e.g., their knowledge, skills, and attitudes, as well as demographic diversity. Culture addresses the beliefs held by team members about their Composition and Context, among other variables (Salas et al, 2015). These C's may not able to be influenced by an ITS, but an ITS that can take them into account in its feedback decisions would likely be better than one that does not.

Meanwhile, the six core processes and emergent states focus on team dynamics and include: Cooperation, Coordination, Cognition, Conflict, Coaching, and Communication (Salas et al, 2015). Cooperation consists of the approaches, behavioral actions, and principles of the team members that lead to the members functioning as a single team, a reflection of team skills. Relatedly, Coordination involves using task-related skills and team skills to reach desired outcomes. Communication is closely tied to these first two considerations as teammates establish and change their approaches and principles. A team's Cognition is defined as a shared understanding of what the team needs to accomplish as well as how it will be accomplished. Cognition can be deeply influenced by Communication, as it is developed through interactions between the members of a team. Coaching, or leadership, is also an important factor in determining the outcome of a team's efforts. Lastly, Conflict, which can be either relational or task-based, is generally considered bad for teams, but small amounts of conflict under the right conditions can actually be beneficial (Salas et al, 2015). Conflict can be defined as a perception of incompatible interests, approaches or views among at least two members of a team (Salas et al, 2015).

Behavioral Markers

A behavioral marker is a concrete, measurable indicator of a more abstract construct such as Cooperation. For example, behavioral markers of the more abstract concept of "embarrassed" might include eyes looking down and hands covering the face. Behavioral markers, if they can be validated, serve as a much more reliable indicator of team performance than traditional self-reported Likert scales. Sottilare et al. (in press) have attempted to validate behavioral markers for several of the nine C's via a meta-analysis of the team literature. For example, the number of times that team members verbally acknowledge when a team member shares information is a positive marker for Communication. On the other hand, the frequency of requests for clarification is a negative marker for Communication.

When interpreting and implementing the nine C's for teamwork, the authors first looked to the above effort as well as to the team assessment literature to find validated behavioral markers to evaluate the C's within their team tutoring system. When implementing one of the C's for which no behavioral markers have been identified, the authors chose different, but related strategies. These strategies are discussed in the three case studies which follow.

CASE STUDIES

The first case study describes a team tutor for small teams (dyads) in which the team members have identical roles. The second and third case studies explore teams which expand beyond two members with different roles.

Case Study 1: The Surveillance Task

The Surveillance Task was the initial case study to determine the attributes of an effective team tutor. It has been previously described in more detail as the Reconnaissance Task (Bonner et al. 2016; Bonner et al. 2015). The Surveillance task serves as simple proof-of-concept and testbed for evaluating a team-based intelligent tutoring system for training soldiers on a military task. Additionally, the task was designed to be able to support assessment of individual and team feedback types provided by the intelligent tutoring system. Team feedback refers to feedback which is given to both team members to address the overall performance of the team, while individual feedback is given to a specific team member based on his or her performance. Included in this initial case study are three of the C's: Communication, Coordination, and Context.

Task description. The task consists of a team of two distributed individuals located in separate offices (see Figure 1). Each participant has a desktop computer with the scenario running in VBS2 (a military game engine), an ITS on the same computer developed using GIFT, an intercom system for communication, and headsets for scenario audio. Each participant controls a virtual avatar placed on the roof of a building within the virtual environment. Each player is responsible for one half of a 360-degree environment (180-degree sectors) where opposing force (OPFOR) avatars appear at random, moving from one participant's sector to the other. To keep track of all OPFOR, participants must continuously review or scan their sector.

Two participants work together to complete the five-minute surveillance task for four trials. To successfully complete this task, they must also complete three subtasks: Transfer, Acknowledge, and Identify. For Transfer, a participant warns the teammate when an OPFOR is headed to the teammate's sector, both verbally and with a keystroke. Immediately following a transfer notification, the teammate who receives the notification must immediately issue an acknowledge message with a verbal response and keystroke. Finally, once the OPFOR has actually moved across the boundary and been spotted by the teammate in that new sector, that teammate identifies the OPFOR with a keystroke. While this overall task may seem simple, as the number of OPFOR increases, players become quite engaged, and performance is by no means perfect.

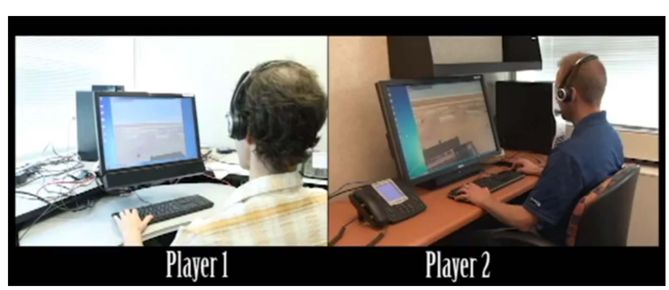

Figure 1: Task 1 Participant Environment

Task Metrics. In assessing the surveillance task, participants are evaluated as individuals and as teams. They are individually assessed on the accuracy and timeliness of their OPFOR transfer, their acknowledgments, and their identifications. As a team, they are assessed on the number of correct coordinated behaviors by both players. We utilize the C's for analysis of team performance. See Table 1.

The team's Coordination is determined by the total number of transfers which have matching acknowledges (transfer-acknowledge pairs) divided by the total number of OPFOR who crossed sectors boundaries. This leaves us with a ratio or percentage for each of the four five-minute trials. Communication is measured as the count of communications by the team members in audio files recorded from each session. Context is determined by assessing the content of the communications. Each communication is given a qualitative categorization code, including task-related (uttered as part of the transfer, acknowledge, or identify tasks), critique (e.g., one player suggests a behavior to improve or help performance), and conversational, for example. This behavioral coding approach is similar to the scheme used in Chorney, McMurtry, Chambers, & Bakeman (2015). Note that behavioral coding initially requires a human in the loop, which removes the possibility of real-time ITS feedback on Context; however, if the tutor were to be used at scale, a machine learning system could be trained to do this coding.

As the goal of the project was to determine how to implement ITSs for teams, we wanted to make sure we were appropriately assessing teams. Coordination specifically was implemented in assessing the relation of two co-dependent tasks. Thus, we can see how consistently the team acted throughout each trial. By including the acknowledge function, we are not only creating an authentic military task (Author et. al, 2015), we are also able to ensure that teamwork is reviewable. For communication, it was plausible to monitor the rate of communications. Previous work from Lamb, Lamb, Stevens & Caras (2014) suggests importance of classifying communications in teams. In some teams, little verbal interaction is needed, while in others the opposite is true.

Table 1: C's for the Surveillance Task

The C	Markers	Data
Coordination	Transfer-Acknowledge pairs	Logs of Transfer & Acknowledge keystrokes and OPFOR crossings
Communication	Number of utterances	Analysis of audio recordings
Context	Category of utterances	Behavioral coding of recorded audio

Case Study 2: Search and Destroy (SD)

The next case study which sought to implement the 9C's was the Search and Destroy Task (SD). The purpose of this task was to begin evaluating an ITS's ability to improve team performance on a three-person team in which participants have different roles. SD sought to move the project to Unity3D while still incorporating GIFT. Additionally, in utilizing Unity, we would be able to build upon the new testbed with technology such as head mounted displays, non-keyboard centric controls, and the potential for further expansion as innovation is developed. When implementing this task, the authors included measures based on Communication, Coordination, and Cooperation. While the scenario has been developed for this task and the ITS planned, the ITS has not yet been implemented.

Task description. SD is a three-person task which takes place in a simple virtual environment (see Figure 2). In this task, two team members control an Apache Helicopter (Pilot and Gunner) while the third team member serves as an aerial intelligence officer (AIO). The Pilot flies the helicopter, the Gunner destroys targets, and the AIO directs the helicopter based on multiple views of the environment. The main objective of the task is to eliminate as many OPFORs as possible.

This task was designed specifically with the goal of high interdependence among the players. For example, if the Pilot does not fly steadily, the Gunner cannot accurately aim the gun at the target. If the AIO does not communicate smoothly with the Pilot, time is wasted when preparing for the next target. The team members use scripted communication to accomplish their mission.

The participants' working environment is akin to that of the Surveillance Task, with each member sitting in separate offices with a desktop computer, headsets, and an intercom for communication. In addition to keyboards, the pilot and gunner were provided with a game controller to maneuver the helicopter and mounted weapon respectively.

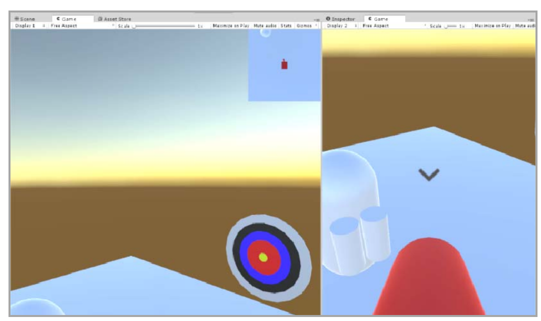

Figure 2: Search and Destroy Environment

Task Metrics. While the SD task involving multiple metrics, several drew on the C's, with Communication being the main focus (see Figure 3). The Pilot's communications are to request shot steadiness from the gunner and confirm potential enemies from the AIO. The Gunner issues positioning requests to the Pilot and informs the AIO when targets are destroyed. The AIO informs the helicopter team (Pilot and Gunner) of obstacles and enemies. Each communication must be acknowledged. To assess communication effectiveness, we count the percentage of communications acknowledged and their timeliness with the events within the scenario. See Table 2. Note that while the C's of Communication and Coordination are present in both Tables 1 and 2, they are defined differently based on the task.

A key element of successful team performance in the SD task is backup behavior (Porter et al., 2003), when one teammate notices that another has missed an obligation and covers. E.g., if the Gunner tells the Pilot and AIO about a newly spotted target before the AIO notifies them. Backup behavior is part of Cooperation. Cooperation in this context can be measured as the ratio of targets communicated by the AIO vs. the total number of targets engaged. Coordination is a measure of successful Pilot-Gunner interactions, and can be measured by the ratio of time that the gun can feasibly engage the target per game engine calculations, vs. the total time in which the Gunner requested steadiness from the Pilot.

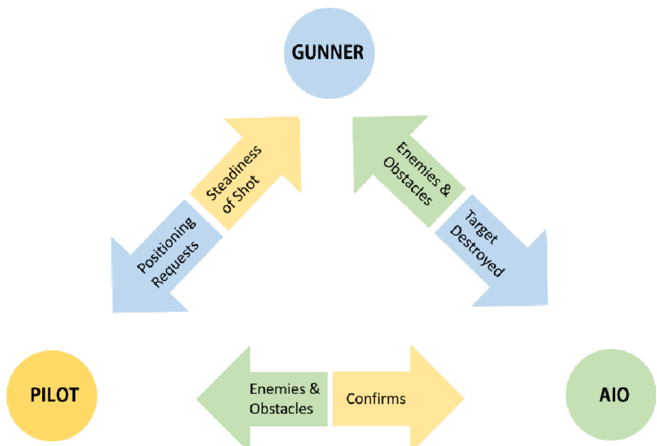

Figure 3: Search and Destroy Communication Flow Chart

Table 2: C's for the SD Task

The C	Markers	Data
Communication	% comms. acknowledged	Comm keystrokes; coded audio recordings
	Latency of comms. after trigger events	Timing of comm. keystrokes, timing of scenario events
Cooperation	% total targets engaged that were noted by AIO	Target identification keystrokes; coded audio recordings
Coordination	% of gun targeting time that the helicopter was held in range	Trig. calculation of "in range"; comm. keystrokes; coded audio recordings.

Case Study 3: Surveillance with Sniper (SwS)

The second iteration of the team surveillance task explores the requirements for an effective intelligent tutor for a group of two or more people with different roles. Specifically, this task seeks to (1) test the potential impact of an ITS for a team of three people, (2) examine the effect of public versus private feedback on the ease with which a player can learn a new role, and (3) examine the effect of cross-role experience on team cognition.

Task description. SwS adds a new role to the two-person task. The two original roles (here called Spotters) still exist and have roughly the same tasks; however, SwS differs from the Surveillance Task in that it contains three types of targets: civilians, OPFORs carrying weapons, and OPFORs carrying weapons and wearing IED vests.

The two Spotters are additionally responsible for transferring these targets to the third member of their team, the Sniper. The following example details more precisely the sequence and type of events and actions involved in this task.

1. Spotter 1 sees an entity in her zone heading towards the boundary and Transfers to Spotter 2.
2. Spotter 2 Acknowledge the transfer.
3. Spotter 2 Identifies new entity present in sector for Sniper to inspect.
4. Sniper Acknowledges the identify.
5. Sniper Assesses Threat using binoculars: Civilian, OPFOR Level 1 or OPFOR Level 2.

An additional change implemented in this study comes in the experimental design. First, each member in the team is cross-trained on both the Spotter and the Sniper role via a tutorial video which asks players to practice the keystrokes and guides them through the tasks they will be completing in the trials which follow. Then, for the first three trials, participants complete the team task in a single role, and in Trial 4, two people switch roles. This approach enables a comparison of a performance of a first-time Spotter (or Sniper) in Trial 1 with a first-time Spotter (or Sniper) in Trial 4, where a participant is inexperienced in the specific role but familiar with the training scenario and its general dynamics.

The other independent variable for this experiment is feedback type, of which there are two types: private and public. The contents of the feedback are similar, focusing on individual actions, but private feedback is given only to individuals, while public feedback is displayed to all team members.

Task metrics. In the SwS, data on five C's are collected: Cognition, Coordination, Communication, Cooperation, and Conflict. Behavioral markers for each of these C's were integrated into the experimental design (see Table 3 for the markers that are new in the SwS task that go beyond the Surveillance Task). Additionally, by introducing a task switch for two of the three players and initial cross-training for all players in SwS, the researchers manipulated the amount of shared understanding, or cognition, within the team.

The Sniper is individually evaluated on timing (speed to acknowledge, speed to Assess Threat) and accuracy of Assess Threat. The team is still evaluated on its Coordination by its performance for transfer-acknowledge pairs, and is additionally evaluated on this C by the total number of identifies which have matching and correct threat assessments divided by the total number of targets which could have been identified. This ratio exists for each of four five-minute trials. To measure Communication, Sociometric badges produced at MIT (Olguín et al., 2009) are used to measure speaking time amount and speech patterns, which are used to identify interruptions. Cognition, which is manipulated in this experiment, is measured using a post-study role knowledge test, while Conflict is measured via qualitative analysis of subjective reports of role-purpose. This analysis could be done manually or with a rule-based ITS system such as ConceptGrid (Blessing, Devasani, Gilbert, & Sinapov, 2015), since there is linguistic structure in the responses. Lastly, Cooperation is measured via the behavioral marker of collective efficacy, or belief in one's team's ability to perform in team-specific tasks. The scale used to collect this self-reported data is based off recommendations by Bandura (2006).

Table 3: C's for the SwS Task that supplement the C's already present from the Surveillance Task

The C	Markers	Data
Communication	Speaking time	Sociometric time output
	Number of interruptions or overlapping comms.	Sociometric speech data output
Cognition	Knowledge of Spotter goals, Sniper goals, and task flow	Score out of 34 on a post-study test
Cooperation	Collective Efficacy	5-question subjective survey responses
Coordination	Identify-Assess Threat pairs	Logs of Identify & Assess Threat keystrokes and target crossings
Conflict	Alignment of task- and role-related beliefs and assumptions	Coded similarity of 3 post-session survey question responses

SUMMARY

Three case studies of intelligent tutors for teams have been described in an effort to illustrate an approach to integrating the nine C's of teamwork as performance measures. In the Surveillance Task, the goal was to create a task for assessing individuals on a simple two-person task. We leveraged Communication, Coordination, and Context. These proved to be useful in teams with the same roles. Specifically, Cooperation was measured via Transfer-Acknowledge pairings. From this we were able to build upon interaction to have the team variable Identify as a marker for a team's communication and context. When utilizing the C's again in the SD, it was more difficult to incorporate them due to the task's goal of training for multiple roles. Successful interdependence of the Pilot, Gunner, and AIO relied on messages being received in the right sequence; coordination was dependent upon communication. Cooperation could be measured from backup behavior. SwS utilized the structure of the Surveillance Task's implementation of cooperation as a foundational C. In doing so, more C's were able to be incorporated as action dictated communication.

The main conclusion in operationalizing C's is the process of selecting C's and defining metrics. As shown in Table 1, Table 2, and Table 3, we determined which behavioral markers to use for each C, and what data to use to measure that behavioral marker. It is worth noting that these case studies, most of the C's were measured after the fact rather than in real-time. This approach is useful for an after-action review form of team feedback by the team tutor. In future work, however, as technology for measuring communications and human interactions increases in its resolution, the authors look forward to measuring C's in real-time and giving feedback appropriately. The authors hope that this process will serve as a model for future authors of team tutoring systems.

ACKNOWLEDGMENTS

The work was completed with support from a U.S. Army Research Laboratory – Human Research & Engineering Directorate (ARL-HRED). Statements and opinions expressed in this text do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES

Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: Example-tracing tutors. *International Journal of Artificial Intelligence in Education, 19*(2), 105-154.

Bandura, A. (2006). Guide for Constructing Self-Efficacy Scales. In F. Pajares & T. Urdan (Eds.), *Self-Efficacy Beliefs of Adolescents* (307-337). USA: Information Age Publishing.

Blessing, S., Devasani, S., Gilbert, S., & Sinapov, J. (2015). Using ConceptGrid as an Easy Authoring Technique to Check Natural Language Responses. *International Journal of Learning Technology, 10*(1), 50-70.

Bonner, D., Gilbert, S., Dorneich, M. C., Winer, E., Sinatra, A. M., Slavina, A., MacAllister, A., Holub, J. (2016). *The Challenges of Building Intelligent Tutoring Systems for Teams.* Paper presented at the Human Factors & Ergonomics Society (HFES) Annual Meeting, Washington, D.C.

Bonner, D., Walton, J., Dorneich, M. C., Gilbert, S. B., Winer, E., & Sottilare, R. A. (2015). The Development of a Testbed to Assess an Intelligent Tutoring System for Teams. In *Proceedings of the Workshops at AIED 2015*. Madrid, Spain.

Chorney, J. M., McMurtry, C. M., Chambers, C. T., & Bakeman, R. (2015). Developing and Modifying Behavioral Coding Schemes in Pediatric Psychology: A Practical Guide. *Journal of Pediatric Psychology*, *40*(1), 154–164. http://doi.org/10.1093/jpepsy/jsu099

Lamb, C., Lamb, J., Stevens, R., & Caras, A. (2014, June). Team Behaviors and Cognitive Cohesion in Complex Situations. In International Conference on Augmented Cognition (pp. 136-147). Springer International Publishing.

Olguín, D. O., Waber, B. N., Kim, T., Mohan, A., Ara, K., & Pentland, A. (2009). Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior. *IEEE Transactions on Systems, Man, and Cybernetcis-Part B: Cybernetics, 39*(1), 43-55. http://doi.org/ 10.1109/tsmcb.2008.2006638

Porter, C. O., Hollenbeck, J. R., Ilgen, D. R., Ellis, A. P., West, B. J., & Moon, H. (2003). Backing up behaviors in teams: the role of personality and legitimacy of need. *Journal of Applied Psychology, 88*(3), 391.

Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2015). Understanding and improving teamwork in organizations: A scientifically based practical guide. *Human Resource Management*, *54*(4), 599-622.

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). The Generalized Intelligent Framework for Tutoring (GIFT). Concept paper released as part of GIFT software documentation. Orlando, FL: US Army Research Laboratory–Human Research & Engineering Directorate (ARL-HRED).

Sottilare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J., & Gilbert, S. B. (in press). Designing Adaptive Instruction for Teams: A Meta-Analysis. *International Journal for Artificial Intelligence in Education*.

Walton, J., Dorneich, M. C., Gilbert, S., Bonner, D., Winer, E., & Ray, C. (2014). Modality and timing of team feedback: Implications for GIFT. In *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym2)* (pp. 199-207).